Musical Muscle Memory and the Haptic Display of Performance Nuance

Chris Chafe
cc@ccrma.stanford.edu

Sile O'Modhrain
sile@ccrma.stanford.edu

Center for Computer Research in Music and Acoustics, Music Department
Stanford University

Abstract

We have begun exploring extraction and editing of nuances of a performance through the sense of touch. Expressive variations in MIDI piano recordings were obtained, limiting the initial study to timing and velocity information. A force-feedback interface displays in real time an analysis of the performer's musical conception and can be used to graft aspects of one performance onto another.

1 Introduction

A challenging analysis problem has haunted one of the authors for years, usually mentioned in terms of how synthesis could benefit from a deeper understanding of performance. Posed as conjecture, it's to imagine if two string quartets were to perform the same piece on different nights: the first night's performance is competent, and the audience is happy enough about it. The second night the performance is simply stunning, transcendent, and the audience leaves ecstatic. Part of the problem poses the question of imagining the differences in terms of quantities which would be acoustically-measurable differences between the performances. A second, possibly more difficult part of the problem, is in comprehending such a wealth of detail so that the analysis is imageable and useful.

A second interest motivating this study is to further exploit the sense of touch in music editing tasks. Beyond automated mixer controls, digital editing involves only display to the eye and ear. However, in the physical creation of music, sounding events are registered by the hand and ear [Chafe, 1993] [Gillespie, 1995]. Present digital technology can be adapted to incorporate the kinesthetic (muscular), tactile, and vibro-tactile (cutaneous) senses, modalities well-suited for data that depicts time and motion.

Performances of the same music can have vastly different feelings even when constrained by a fully-notated score. For simplicity, a short piano excerpt was chosen for this study and independent renditions were compared in terms of event timings and key velocities. As listeners, we are acutely sensitive to these differences, but it is more likely that we are only aware of their aggregate effect, for instance, the feeling that one passage was played more forcefully than another. What are the note-level differences, how are they structured, and are such structures the basis for the affect?

The hope that differences of affect can be characterized and displayed leads to the further possibility of manipulating recorded or synthesized performances. A computer-controlled force-feedback interface was programmed to display aspects of performance and manipulate them in real time. Haptic display has the advantage of communicating directly to the motor senses, the same that are involved in musical performance. The word "haptic" is employed to describe devices that engage both the kinesthetic and tactile senses. In our work, the quantities displayed to the observer are ideally a replay or recasting of human motor commands which might have created or accompanied a performance. The end-result is a prototype system that allows the observer to feel musical *feeling* through the real-time display of parameters analyzed from performance. Because the controller permits direct interaction with its display, the performance can be edited in an intuitive manner.

2 Method

An excerpt from the opening of Beethoven's Piano Sonata, Opus 109, was recorded by two excellent pianists using a Yamaha Disklavier grand piano, Figure 1. Recorded data was transferred into standard MIDI file format and analyzed in several steps (with the Stella programming environment, a Lisp package for symbolic musical manipulation [Taube, 1993]). First, the two performances were

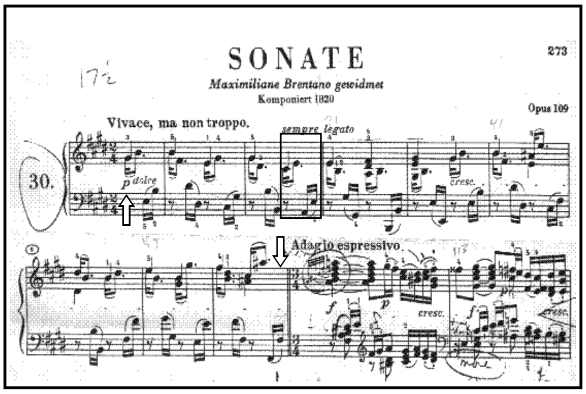

Figure 1: *Two performances of the opening of Beethoven's Piano Sonata, Opus 109, were recorded by Yamaha Disklavier. Note timings and key velocity data were transferred to standard MIDI files.*

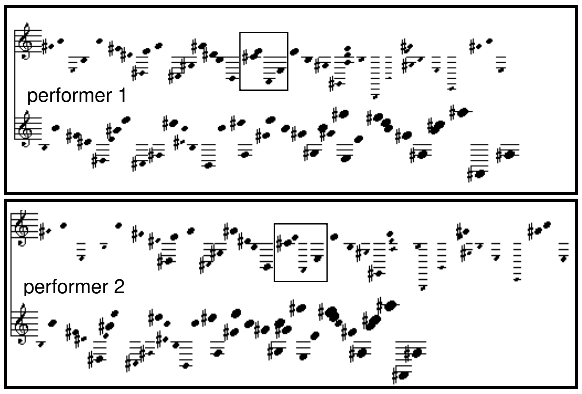

Figure 2: *Distinct short-term shapes are found in raw data displayed from the first 77 notes (marked by arrows in Figure 1). Note placement is proportional to time and size is proportional to key velocity.*

matched up in terms of detected pitches. Our performers were not supervised in any way and were free to submit what they wished. Approximately 2% of the notes did not match up for a variety of reasons, including wrong notes and order differences in chords. Since our project is ultimately directed at acoustically recorded performances, and we expect an even greater error rate in the transcription process, this level of mismatch was acceptable [Chafe and Jaffe, 1986]. A matching algorithm was applied, working from the beginning of the data and pairing equivalent pitches between the two performances. Discrepancies were eliminated and the resulting data set of matched pitches provided the basis for initial experimentation.

2.1 The Moose

Performance data was transferred to a program written in C++ commanding a MIDI synthesizer and the moose, a two-dimensional haptic display device. The moose is essentially a powered mouse-like pointing device. It consists of a puck or manipulandum in the center coupled to two linear voice-coil motors through two perpendicularly oriented flexures. The double flexures conveniently decouple the 2-axis motion of the puck into two single-axis motions at the linear motors. The puck's motion is restricted to an area in the horizontal plane approximately 3 inches on a side.

The moose was designed as part of a larger project based at CSLI, Stanford, to investigate the possibility of using haptic technology to display elements of graphical user interfaces such as window edges, buttons, etc. to blind computer users [O'Modhrain, 1995]. The prototype display has proven the feasibility of the approach, and will continue to be developed alongside our exploration into the use of haptics as a component of a digital music editing systems.

2.2 First Results

Restricting the data to the first eight and a half measures of the Beethoven focused initial analysis on a passage consisting only of running sixteenth-note rhythms. For further simplification, pedal information and durations were ignored. The collected note onsets and key velocities show short-term shapes superimposed on longer-term phrasings. The moose was programmed to directly display key velocity data in the form of an elastic wall. The observer presses the puck to a virtual wall whose stiffness depends on the MIDI velocity being sent to a piano synthesizer. While the performance is sounding, the wall portrays a strong sense of note-to-note variation. As can be seen in Figure 2, some of the note-to-note instantaneous changes are quite abrupt, and a modification was made to display a small mixture of instantaneous key velocity plus a moving average of key velocity whose window is centered on the current note. A rather satisfactory sensation of dynamic phrasing results.

The next refinement consisted of combining onset timings with velocity data to establish an abstract *effort* parameter. Effort, in this sense, represents the directions a conductor might impart to an orchestra. High effort corresponds to faster & louder, low effort to relaxed & softer. However, *ritardando* & *crescendo* can also elicit strong effort, as in the end of the passage studied. A formula to represent these relationships was devised (based on the simplification that the score excerpt only consists of sixteenth notes, which are nominally 125 msec):

$$effort = nv * (1/r + C * r^2)$$

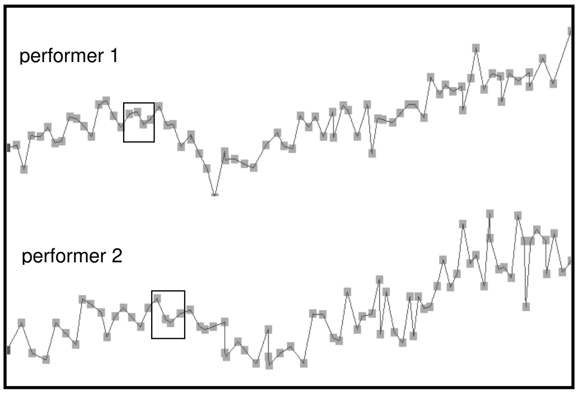

Figure 3: Effort *vs. time is compared for the same passage as Figure 2. The effort quantity is derived from note onset timings and key velocity. Total duration has been normalized for ease of comparison.*

nv is normalized velocity scaled from 0.0 to 1.0 from the recorded range of velocities, r represents the time interval from the onset of the previous note, and C is a coefficent to bring the nominal rhythm value into range.

Figure 3 shows a graph of effort derived for the same passage as Figure 2. Multi-measure swells correspond to long-term phrasing. Short-term shapes can be seen in note groupings of 2 - 6 notes at a time. The two performances have the keenest difference on this short-term time-scale. Groupings are sometimes similar but shapes are distinct. For example at note 17, a four-note groupin appears (marked by boxes in the figures). Through the effect of a single note, the shape differs between the two performances.

2.3 Manipulations and Muscle Memory

The moose displays the two time scales as separate sensations: a background long-term motion and superimposed, faster foreground shapes. In the background, long-term changes are displayed by averaging the effort parameter with a moving window and causing the virtual wall position to change smoothly. In the foreground, instantaneous effort values affect the wall's compliancy, with higher effort values causing a stiffer spring, Figure 4. The observer quickly trains on differences between the two performances.

A third performance can be created as a product of the first two through linear interpolation of onset rhythms and velocities. The wall's length is used as the interpolation control. At the wall ends, the observer experiences one performance or the other and, in between, an interpolated version. Sliding along the wall in real time allows grafting of one performance to the other.

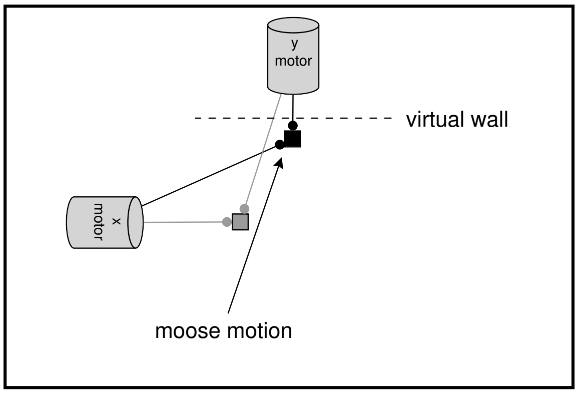

Figure 4: *The moose, a powered mouse, consists of two linear voice-coil motors controlling the location of a puck. Virtual objects and surfaces are displayed by force-feedback. The performance analysis is displayed by changes to a virtual wall's location and compliance in real time while the music is played.*

The prototype system suggests that in-real-time experiences through haptic devices such as the virtual wall can be coupled with sound to offer a rich display for performance analysis and editing. Imaging and memory of patterns is enhanced by appealing to muscle memory. One way to imagine this is to contrast the method with an out-of-time graphical display, such as in Figure 3, or a static haptic display which would project Figure 3 onto a touchable surface. Spatial displays excel for side-by-side pattern discrimination, and performance shapes, such as those briefly discussed, are easily found. Animated spatial displays increase dimensionality, often to include time. The force-feedback system is used to go the other way, to reduce the data into one simplified, intuitive, musical dimension such as the effort parameter. The observer is able to experience, vicariously, the performer's own feeling of effort during performance.

3 Summary: Haptics and Sound

Haptic perception *of the signal* has been lost through changes in music-making technology. The mechanical musical world consists of direct manipulation of sound-producing mechanisms and a sense of their vibration. The analog world replaced this with the feel of various specific control devices or the feel of motion of the recording medium. The digital world has reduced this further to a few general purpose controllers and displays, eg. mouse, keyboard, CRT.

This study has already shown us that there are indeed parameters within music which can be manipulated to allow a performer, composer or music editor to traverse the space between two totally different interpretations of the same piece. We have demonstrated that we can make these parameters apparent to the kinesthetic and vibrotactile senses, those same senses which, in live performance, complete the musician's feed-back loop. With a few simple haptic interface tools we can bring back to the editing process some of its former intuitiveness and flexibility [O'Modhrain, 1995].

Specifcally, what we have lost in the transition to mouse-based digital music editing environments is the close contact which sound engineers once enjoyed with their media. We can design new haptic controls and program their "feel" by making them more or less resistant to being moved. A shuttle wheel detent is, for example, easily mediated by motors. And a detent could represent variously manipulator or signal state.

Unique physical operations on sound persist today as metaphors in digital audio editing tools. For example, records are scratched back and forth, tapes are slowed and sped up. The musical arts themselves are strongly influenced by such technologies which often form a basis for new genres of technologically-influenced music. We look forward to enjoying the artistic output inspired by the programmable, multi-modal, and physically coupled interfaces of the future which will feature haptic components.

The authors gratefully acknowledge contributions to the project from our colleagues George Barth, Brent Gillespie, Craig Sapp, and Frederick Weldy. The Archimedes Project at Stanford University's Center for Study of Language and Information provides ongoing support for development of haptic access to graphical user interfaces.

References

[Chafe, 1993] Chris Chafe. Tactile Audio Feedback. *Proceedings of the ICMC, Tokyo*, 1993.

[Taube, 1993] Heinrich Taube. Stella: Persistent Score Representation and Score Editing in Common Music. *Computer Music Journal,* **17**(4), 1993.

[Chafe and Jaffe, 1986] Chris Chafe and David Jaffe. Source Separation and Note Identification in Polyphonic Music. *Proceedings of the IEEE Conference on Acoustics Speech and Signal Processing, Tokyo* (2): pp. 25.6.1-25.6.4, 1986.

[Gillespie, 1995] Brent Gillespie. Haptic Display Of Systems With Changing Kinematic Contraints: The Virtual Piano Action. *Dissertation, Dept. of Mechanical Engineering, available as Stanford Music Department Report STAN-M-92* Stanford University, 1995.

[O'Modhrain, 1995] Sile O'Modhrain. The Moose: A Haptic User Interface For Blind Persons With Application to the Digital Sound Studio. *Stanford Music Department Report STAN-M-95* Stanford University, 1995.